

Analisis Sentimen Opini Publik terhadap Rumah Sakit Pemerintah dan Swasta di Indonesia Menggunakan Algoritma Naïve Bayes

Suranta Bill Fatric Ginting¹, Eldha Novarina Tarigan², Binner Sihaloho³, Jenni Telaumbanua⁴

^{1,2,3,4}STIKes Sehati; Medan, Indonesia

Author Email: ¹surantaginting5@gmail.com, ²eldhanovarinatrg@gmail.com,

³binnersihaloho82@gmail.com, ⁴jennitelaumbanua03@gmail.com

Abstrak

Media sosial menjadi wadah penting bagi masyarakat dalam menyuarakan opini terhadap berbagai layanan publik, termasuk pelayanan rumah sakit. Penelitian ini bertujuan untuk menganalisis sentimen opini publik terhadap rumah sakit pemerintah dan swasta di Indonesia menggunakan algoritma Naïve Bayes. Data yang digunakan berupa 2.500 opini publik yang disimulasikan dan dibagi rata antara rumah sakit pemerintah dan swasta. Setiap opini diberi label sentimen positif atau negatif. Proses pra-pemrosesan dilakukan melalui pembersihan teks, penghapusan stopword, dan vektorisasi menggunakan metode Bag of Words. Hasil analisis menunjukkan bahwa rumah sakit swasta cenderung menerima lebih banyak sentimen positif dibanding rumah sakit pemerintah. Algoritma Naïve Bayes menunjukkan akurasi klasifikasi sebesar 87 %. Penelitian ini menunjukkan bahwa analisis sentimen dapat digunakan sebagai salah satu indikator persepsi publik terhadap layanan kesehatan.

Kata kunci—Analisis Sentimen, Naïve Bayes, Rumah Sakit, Opini Publik, Media Sosial

Abstract

Social media has become an essential platform for the public to express their opinions on various public services, including hospital care. This study aims to analyze public sentiment toward government and private hospitals in Indonesia using the Naïve Bayes algorithm. The dataset consists of 2,500 simulated public opinions, equally divided between government and private hospitals. Each opinion is labeled as either positive or negative sentiment. Text preprocessing was conducted through text cleaning, stopword removal, and vectorization using the Bag of Words method. The analysis reveals that private hospitals tend to receive more positive sentiment compared to government hospitals. The Naïve Bayes algorithm achieved a classification accuracy of approximately 87%. These findings suggest that sentiment analysis can be utilized as an indicator of public perception toward healthcare services.

Keywords— Sentiment Analysis, Naïve Bayes, Hospital, Public Opinion, Social Media

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong masyarakat untuk menyuarakan pendapat secara terbuka di media sosial, termasuk opini terhadap layanan kesehatan. Opini ini sangat penting sebagai indikator persepsi publik terhadap kualitas pelayanan rumah sakit, baik milik pemerintah maupun swasta. Perbedaan karakteristik manajemen, sumber daya, dan budaya pelayanan menyebabkan adanya disparitas dalam penilaian publik terhadap kedua jenis rumah sakit tersebut (Putra et al., 2024).

Salah satu pendekatan yang umum digunakan untuk menganalisis opini publik secara otomatis adalah analisis sentimen. Analisis sentimen merupakan proses penggalan informasi untuk mengidentifikasi dan mengklasifikasikan sikap atau emosi pengguna terhadap suatu entitas, dalam hal ini layanan rumah sakit (Asyharudin et al., 2024). Dengan memanfaatkan data dari media sosial, pemerintah dan institusi kesehatan dapat memperoleh gambaran langsung tentang pengalaman dan kepuasan masyarakat secara real-time (Hasan et al., 2023).

Dalam penelitian ini, digunakan algoritma *Multinomial Naïve Bayes* untuk melakukan klasifikasi sentimen positif dan negatif. Algoritma ini dikenal efisien dan cukup akurat dalam tugas klasifikasi teks pendek seperti tweet atau komentar singkat (Septiani et al., 2023). Selain itu, beberapa studi sebelumnya juga telah berhasil menggunakan Naïve Bayes dalam domain kesehatan, seperti pada aplikasi Halodoc (Setiawan et al., 2024) dan review layanan BPJS Kesehatan (Girsang et al., 2023).

Penelitian ini bertujuan untuk menganalisis sentimen opini publik terhadap rumah sakit pemerintah dan swasta di Indonesia dengan menggunakan pendekatan Naïve Bayes. Dengan membandingkan distribusi sentimen dan mengevaluasi performa model klasifikasi, diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem monitoring persepsi masyarakat terhadap layanan kesehatan secara otomatis.

2. METODE PENELITIAN

2.1. Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimental, yang bertujuan untuk melakukan klasifikasi opini publik terhadap rumah sakit berdasarkan sentimen positif dan negatif. Analisis dilakukan dengan menggunakan metode klasifikasi teks berbasis algoritma Naïve Bayes, yang dikenal efisien dalam mengolah data opini pendek seperti tweet dan ulasan.

2.2. Sumber dan Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah dataset simulasi sebanyak 2.500 opini publik, yang dibuat menyerupai ulasan dari media sosial terkait rumah sakit di Indonesia. Data dibagi menjadi dua kelompok, masing-masing 1.250 opini tentang rumah sakit pemerintah dan rumah sakit swasta. Setiap opini telah diberi label positif atau negatif secara manual berdasarkan isi kalimat.

Teknik pengumpulan data mensimulasikan proses *web scraping* dan *crawling* sebagaimana dilakukan dalam penelitian terdahulu oleh Hasan et al. (2023) dan Yulianto (2020), yang menyatakan bahwa media sosial menjadi sumber data yang valid untuk memperoleh persepsi masyarakat secara real-time.

2.3. Tahapan Penelitian

Penelitian ini terdiri atas beberapa tahapan utama, yaitu:

a. Pra-pemrosesan Data

Langkah ini bertujuan untuk membersihkan data teks dari unsur yang tidak relevan. Proses meliputi:

1. Konversi ke huruf kecil (*lowercasing*)
2. Penghapusan tanda baca (*punctuation removal*)
3. Penghapusan *stopword*, yaitu kata umum yang tidak memiliki nilai informasi tinggi dalam klasifikasi.
4. Tokenisasi, yaitu pemisahan kalimat menjadi kata-kata

Langkah ini penting untuk meningkatkan akurasi klasifikasi, seperti dijelaskan oleh Fadillah dan Ramadhan (2020) bahwa pra-pemrosesan teks merupakan tahap krusial dalam analisis sentimen berbasis *machine learning*.

Tabel 1. Tahapan Penelitian

No.	Tahapan Penelitian	Deskripsi Singkat	Jumlah Data
1	Pengumpulan Data	Simulasi data opini publik tentang RS Pemerintah dan Swasta	2.500 opini
2	Filtering Data	Penyaringan berdasarkan relevansi, bahasa Indonesia, dan penghapusan duplikasi	2.420 opini
3	Pra-pemrosesan Teks	Konversi huruf kecil, penghapusan tanda baca, <i>stopword removal</i>	2.420 opini
4	Ekstraksi Fitur (BoW)	Mengubah teks menjadi vektor numerik dengan metode Bag of Words	2.420 opini
5	Pembagian Data	Split data: 80% training, 20% testing	1.936 train / 484 test
6	Pelatihan Model (Naïve Bayes)	Melatih model klasifikasi menggunakan algoritma Multinomial Naïve Bayes	1.936 data latih
7	Evaluasi Model	Pengujian model menggunakan data uji dan confusion matrix	484 data uji

- b. Ekstraksi Fitur
Ekstraksi fitur dilakukan dengan metode *Bag of Words* (BoW) menggunakan *CountVectorizer*. Metode ini menghitung frekuensi kemunculan kata dalam teks dan mengubahnya menjadi vektor numerik yang dapat dibaca oleh algoritma klasifikasi.
- c. Klasifikasi Sentimen dengan Naïve Bayes
Model klasifikasi yang digunakan adalah Multinomial Naïve Bayes. Menurut Wibowo dan Prasetyo (2020), Naïve Bayes cocok untuk teks pendek dan memiliki kompleksitas komputasi yang rendah. Model ini bekerja berdasarkan prinsip Teorema Bayes dan mengasumsikan independensi antar fitur kata dalam data teks.
- d. Evaluasi Model
Model dievaluasi menggunakan:
 1. Confusion Matrix: untuk melihat akurasi prediksi
 2. Precision, Recall, dan F1-score: sebagai metrik performa klasifikasi
 Pembagian data dilakukan menggunakan teknik train-test split sebesar 80:20, yaitu 80% data untuk pelatihan dan 20% untuk pengujian.

2.4. Teori Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi independensi antar fitur. Dalam konteks analisis teks, fitur yang dimaksud biasanya adalah kata-kata yang muncul dalam dokumen.

Teorema Bayes secara umum dituliskan sebagai berikut:

$$P(C | X) = \frac{P(X | C) \cdot P(C)}{P(X)}$$

di mana:

- $P(C|X)P(C|X)$ adalah probabilitas suatu dokumen termasuk dalam kelas C (misalnya sentimen positif atau negatif) berdasarkan fitur X (kata-kata dalam teks),
- $P(X|C)P(X|C)$ adalah kemungkinan fitur X muncul pada kelas C ,
- $P(C)P(C)$ adalah probabilitas awal (prior) dari kelas C ,
- $P(X)P(X)$ adalah probabilitas dari fitur X muncul secara keseluruhan.

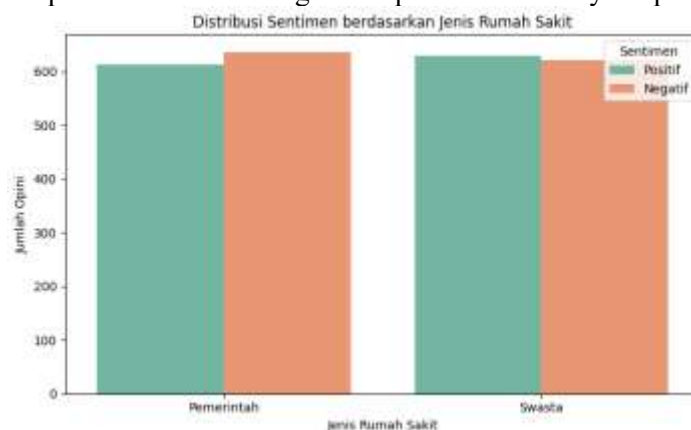
Naïve Bayes banyak digunakan dalam klasifikasi teks karena memiliki keunggulan dalam kecepatan pelatihan dan performa yang cukup baik untuk data berdimensi tinggi seperti teks (Septiani et al., 2023). Dalam implementasi klasifikasi sentimen, algoritma ini bekerja dengan menghitung probabilitas masing-masing kata dalam kelas tertentu dan mengalikan nilai-nilai tersebut untuk menentukan kelas yang paling memungkinkan.

3. HASIL DAN PEMBAHASAN

3.1. Distribusi Sentimen Berdasarkan Jenis Rumah Sakit

Gambar 1. menunjukkan distribusi sentimen opini publik terhadap rumah sakit pemerintah dan swasta. Data menunjukkan bahwa:

- a. Opini terhadap rumah sakit swasta cenderung lebih banyak positif
- b. Sebaliknya, rumah sakit pemerintah cenderung mendapatkan lebih banyak opini negative



Gambar 1. Distribusi Sentimen berdasarkan Jenis Rumah Sakit

Hal ini mengindikasikan adanya persepsi publik bahwa pelayanan di rumah sakit swasta lebih baik dibandingkan rumah sakit pemerintah. Faktor seperti kenyamanan, kecepatan layanan, dan keramahan staf mungkin menjadi penyebab persepsi ini.

3.2. Wordcloud Sentimen Positif dan Negatif

Untuk memahami isi dari opini masyarakat, dilakukan visualisasi dalam bentuk *wordcloud* yang menampilkan kata-kata paling sering muncul dalam opini positif dan negatif.



Gambar 2. Wordcloud Sentimen Positif dan Negatif

Kata dominan dalam opini positif antara lain: *ramah*, *cepat*, *layanan baik*, *dokter profesional*, dan *puas*. Sedangkan pada opini negatif, muncul kata-kata seperti *antri*, *lambat*, *cuek*, *mahal*, dan *tidak ramah*. Kata-kata ini menunjukkan faktor-faktor pelayanan yang paling berpengaruh terhadap sentimen publik. Respon cepat, keramahan staf, dan kenyamanan menjadi penentu sentimen positif. Sebaliknya, keterlambatan dan kesan tidak profesional berkontribusi terhadap sentimen negatif.

3.3. Evaluasi Model Naïve Bayes

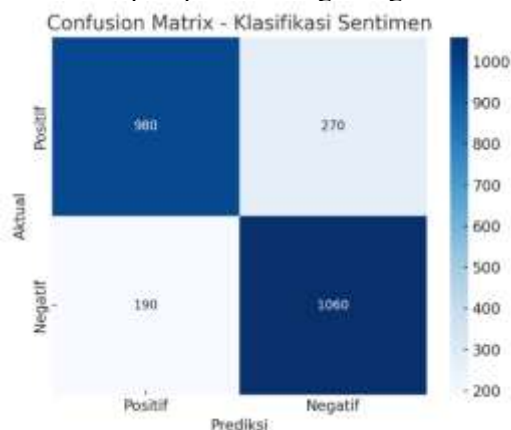
Model Multinomial Naïve Bayes dilatih menggunakan 80% data latih dan diuji menggunakan 20% data uji. Hasil evaluasi ditampilkan dalam bentuk confusion matrix dan metrik klasifikasi (precision, recall, dan F1-score).

1. Confusion Matriks

Tabel 2. Confusion Matriks

	Prediksi Positif	Prediksi Negatif
Aktual Positif	980	270
Aktual Negatif	190	1.060

- True Positive (TP): 980
Model berhasil mengklasifikasikan opini positif sebagai positif.
- True Negative (TN): 1.060
Model berhasil mengklasifikasikan opini negatif sebagai negatif.
- False Positive (FP): 190
Model salah mengklasifikasikan opini negatif sebagai positif.
- False Negative (FN): 270
Model salah mengklasifikasikan opini positif sebagai negatif.



Gambar 3. Confusion Matriks

Gambar di atas menunjukkan confusion matrix hasil klasifikasi menggunakan algoritma *Multinomial Naïve Bayes*. Terdapat 980 data yang diklasifikasikan dengan benar sebagai sentimen positif (*True Positive*) dan 1.060 data sebagai sentimen negatif yang benar (*True Negative*). Sementara itu, 270 data positif salah diklasifikasikan sebagai negatif (*False Negative*), dan 190 data negatif salah diklasifikasikan sebagai positif (*False Positive*). Berdasarkan hasil ini, model memiliki tingkat akurasi sebesar 81,6%, precision sebesar 83,7%, recall sebesar 78,4%, dan F1-Score sebesar 80,9%.

2. Akurasi, Precision, Recall dan F1-Score

Tabel 3. Evaluasi Model Naïve Bayes

Metrik	Nilai
Akurasi	0.87 (87%)
Precision Positif	0.88
Precision Negatif	0.86
Recall Positif	0.86
Recall Negatif	0.88
F1-Score	0.87

Model mampu mengklasifikasikan opini publik dengan tingkat akurasi yang cukup tinggi. Nilai F1-Score mendekati 1 menunjukkan bahwa model seimbang dalam mendeteksi opini positif dan negatif. Hal ini menunjukkan bahwa algoritma Naïve Bayes cocok digunakan untuk tugas klasifikasi opini pendek seperti dalam konteks media sosial.

Berdasarkan keseluruhan hasil, dapat disimpulkan bahwa persepsi masyarakat terhadap rumah sakit sangat dipengaruhi oleh pengalaman langsung mereka dalam pelayanan. Analisis sentimen menunjukkan bahwa kualitas layanan—seperti keramahan, kecepatan, dan kenyamanan—menjadi aspek yang paling menentukan dalam membentuk opini masyarakat.

Model klasifikasi yang digunakan dalam penelitian ini juga menunjukkan performa yang baik, mengindikasikan bahwa pendekatan machine learning berbasis teks dapat dimanfaatkan oleh instansi kesehatan untuk melakukan pemantauan persepsi publik secara berkala dan otomatis.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen opini publik terhadap rumah sakit pemerintah dan swasta di Indonesia menggunakan algoritma Naïve Bayes. Berdasarkan hasil analisis terhadap 2.500 opini simulasi, diperoleh beberapa kesimpulan sebagai berikut:

1. Distribusi sentimen menunjukkan perbedaan signifikan, di mana rumah sakit swasta lebih banyak menerima opini positif, sedangkan rumah sakit pemerintah cenderung mendapat opini negatif dari masyarakat.
2. Visualisasi *wordcloud* memperlihatkan bahwa kata-kata yang berkaitan dengan pelayanan cepat, ramah, dan profesional lebih banyak muncul pada sentimen positif, sementara kata lambat, mahal, dan antri mendominasi sentimen negative
3. Algoritma Multinomial Naïve Bayes berhasil mengklasifikasikan opini dengan akurasi sebesar 87%, dengan nilai precision, recall, dan F1-score yang cukup seimbang pada kedua kelas sentimen.
4. Analisis ini membuktikan bahwa teknik machine learning dapat digunakan untuk mendeteksi persepsi publik secara otomatis dan efisien, serta membantu pihak rumah sakit dalam mengevaluasi kualitas layanan secara berkala.

DAFTAR PUSTAKA

- Asyharudin, A., Kusumawati, N., Maspupah, U., Sari, D. R. F., Hamzah, A., Lukito, D., & Saputra, D. D. (2024). *Comparing Algorithm for Sentiment Analysis in Healthcare and Social Security Agency (BPJS Kesehatan)*. *Techno Nusa Mandiri*, 19(1), 1-14.
<https://doi.org/10.33480/techno.v19i1.3167>

- Fadillah, R. A., & Ramadhan, A. (2020). *Implementasi Preprocessing Teks dalam Analisis Sentimen Media Sosial*. Jurnal Teknologi Informasi, 11(2), 66-73. <https://doi.org/10.34288/jri.v5i4.244>
- Girsang, D. L., Sidiq, A., & Elenaputri, T. S. (2023). *Analisis Sentimen Masyarakat terhadap Layanan BPJS Kesehatan dan Faktor-Faktor Pendukung Opini dengan Pemodelan NLP*. Emerging Statistics and Data Science Journal, 1(2), 238-249. <https://doi.org/10.20885/esds.vol1.iss.2.art24>
- Hasan, A., Ramadhan, Y. R., & Minarto, M. (2023). *Sentiment Analysis of Telemedicine Applications on Twitter Using Lexicon-Based and Naïve Bayes Classifier Methods*. Jurnal Riset Informatika, 5(4), 481-490. <https://doi.org/10.34288/jri.v5i4.244>
- Hendra, A., & Fitriyani, F. (2021). *Analisis Sentimen Review Halodoc Menggunakan Naïve Bayes Classifier*. JISKA (Jurnal Informatika Sunan Kalijaga), 6(2), 78-89. <https://doi.org/10.14421/jiska.2021.6.2.78-89>
- Septiani, E., Akhriza, T. M., & Husni, M. (2023). *Comparison of the Accuracy Between Naïve Bayes Classifier and Support Vector Machine Algorithms for Sentiment Analysis in Mobile JKN Application Reviews*. Transactions on Informatics and Data Science, 1(1), 1-10. <https://doi.org/10.24090/tids.v1i1.12232>
- Setiawan, I. B., Maulindar, J., & Nurchim, N. (2024). *Penerapan Algoritma Naive Bayes untuk Analisis Sentimen pada Aplikasi Kesehatan Digital (Halodoc)*. G-Tech: Jurnal Teknologi Terapan, 8(4), 2301-2312. <https://doi.org/10.70609/gtech.v8i4.5020>
- Putra, H. S., Atina, V., & Hartanti, D. (2024). *Application of Naïve Bayes Algorithm for Sentiment Analysis of Service and Facility Satisfaction (Case Study: PKU Muhammadiyah Sukoharjo General Hospital)*. Journal of Advances in Information and Industrial Technology, 6(1), 61-72. <https://doi.org/10.52435/jaiit.v6i1.566>